

1. Where do I start at NCBI ?

NCBI provides a queryable interface that enables navigation between inter-connected data types and between information for multiple genomes. NCBI brings together the power of large-scale computational analyses and an integrated system of data retrieval with detailed information on maps, sequence, expression, genomes, genes, diseases, phenotypes, publications, protein domains and structures. NCBI provides connections between these disparate types of data by pre-computing related records.

Searches at NCBI can begin in many places. Among these are NCBI's home page (<http://www.ncbi.nlm.nih.gov/>), the Entrez home page (<http://www.ncbi.nlm.nih.gov/Entrez/>), the Zebrafish Genome Resources page (<http://www.ncbi.nlm.nih.gov/genome/guide/zebrafish>) and the *Danio rerio* Taxonomy database home page (<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=7955&lvl=3&srchmode=1>).

1. NCBI's home page

Explore links provided at NCBI's home page to navigate to NCBI's databases, tools and resources. On NCBI's home page, links are provided in the top menu bar, in the left blue bar and under the "Hot Spots" header on the right. From the NCBI home page, submit a direct query against any of the databases included in the pull down menu, including a query against all **Entrez databases**.

Announcements of new and updated resources are provided on the NCBI home page as Highlights in the center of the page and in detail in NCBI's quarterly newsletter. NCBI's newsletter also includes additional announcements, NCBI project descriptions and recent NCBI publications.

NCBI
National Center for Biotechnology Information
National Library of Medicine National Institutes of Health

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

Search for

SITE MAP
Guide to NCBI resources

About NCBI
An introduction for researchers, educators and the public

GenBank
Sequence submission support and software

Literature databases
PubMed, OMIM, Books, and PubMed Central

Molecular databases
Sequences, structures, and taxonomy

Genomic biology
The human genome, whole genomes, and

What does NCBI do?
 Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More...](#)

The Genetic Landscape of Diabetes

Over 17 million Americans have diabetes. Explore the genes discovered thus far with "The Genetic Landscape of Diabetes" as your guide.

Entrez Gene
 You can now use Entrez to search for information centered on the concept of a gene, and connect to many sources of related information both within and outside NCBI.

PubMed Central
An archive of life sciences journals

• **Free fulltext**
 • **Over 200,000 articles from over 140 journals**
 • **Linked to PubMed and fully searchable**
 Use of PubMed Central requires no registration or fee. Access it from any computer with an Internet connection.

NCBI Newsletter

FG++ enhanced NCBI training course

Slots available for FieldGuidePlus Training Course at NCBI

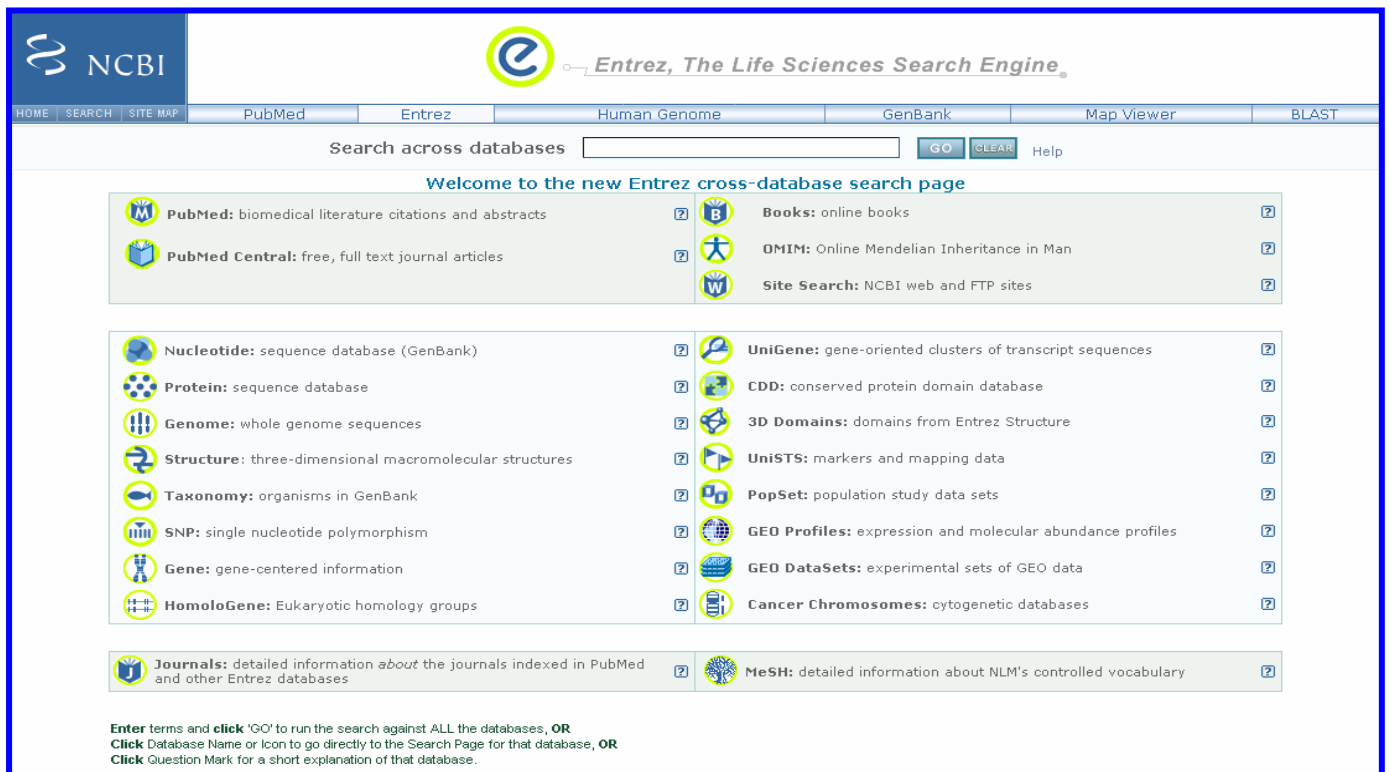
Hot Spots

- ▶ Clusters of orthologous groups
- ▶ Coffee Break, Genes & Disease, NCBI Handbook
- ▶ Electronic PCR
- ▶ Entrez Home
- ▶ Entrez Tools
- ▶ Gene expression omnibus
- ▶ Human genome resources
- ▶ LocusLink
- ▶ Malaria genetics & genomics
- ▶ Map Viewer
- ▶ dbMHC
- ▶ Mouse genome resources
- ▶ ORF finder
- ▶ Rat genome resources

2. Entrez home page

The Entrez home page lists and provides links to all Entrez databases. From the home page, all Entrez databases can be searched simultaneously with a single query. Both simple (single text term) and more complex queries are supported. Complex queries of one or more terms that include boolean operators (AND, OR, NOT) and searches of specified fields. Field qualifiers can be upper or lower case, for example: [organism], [ORGN], [taxonomy_id], [keyword], or [KYWD]. A list of Entrez Search Fields and Qualifiers for each Entrez database is also available at:

(http://www.ncbi.nlm.nih.gov/entrez/query/static/help/Summary_Matrices.html#Search_Fields_and_Qualifiers).



The screenshot shows the NCBI Entrez home page. At the top left is the NCBI logo. To its right is the Entrez logo and the text "Entrez, The Life Sciences Search Engine". Below this is a navigation bar with links: HOME, SEARCH, SITE MAP, PubMed, Entrez, Human Genome, GenBank, Map Viewer, and BLAST. A search bar is located below the navigation bar with the text "Search across databases" and buttons for GO, CLEAR, and Help. The main content area is titled "Welcome to the new Entrez cross-database search page". It features a grid of database icons and descriptions. The databases listed are: PubMed (biomedical literature citations and abstracts), PubMed Central (free, full text journal articles), Books (online books), OMIM (Online Mendelian Inheritance in Man), Site Search (NCBI web and FTP sites), Nucleotide (sequence database (GenBank)), Protein (sequence database), Genome (whole genome sequences), Structure (three-dimensional macromolecular structures), Taxonomy (organisms in GenBank), SNP (single nucleotide polymorphism), Gene (gene-centered information), HomoloGene (Eukaryotic homology groups), UniGene (gene-oriented clusters of transcript sequences), CDD (conserved protein domain database), 3D Domains (domains from Entrez Structure), UniSTS (markers and mapping data), PopSet (population study data sets), GEO Profiles (expression and molecular abundance profiles), GEO DataSets (experimental sets of GEO data), Cancer Chromosomes (cytogenetic databases), Journals (detailed information about the journals indexed in PubMed and other Entrez databases), and MeSH (detailed information about NLM's controlled vocabulary). At the bottom, there is a footer with instructions: "Enter terms and click 'GO' to run the search against ALL the databases, OR Click Database Name or icon to go directly to the Search Page for that database, OR Click Question Mark for a short explanation of that database."

NCBI

Entrez, The Life Sciences Search Engine

HOME | SEARCH | SITE MAP | PubMed | Entrez | Human Genome | GenBank | Map Viewer | BLAST

Search across databases [Help](#)

Welcome to the new Entrez cross-database search page

PubMed: biomedical literature citations and abstracts	Books: online books
PubMed Central: free, full text journal articles	OMIM: Online Mendelian Inheritance in Man
	Site Search: NCBI web and FTP sites
Nucleotide: sequence database (GenBank)	UniGene: gene-oriented clusters of transcript sequences
Protein: sequence database	CDD: conserved protein domain database
Genome: whole genome sequences	3D Domains: domains from Entrez Structure
Structure: three-dimensional macromolecular structures	UniSTS: markers and mapping data
Taxonomy: organisms in GenBank	PopSet: population study data sets
SNP: single nucleotide polymorphism	GEO Profiles: expression and molecular abundance profiles
Gene: gene-centered information	GEO DataSets: experimental sets of GEO data
HomoloGene: Eukaryotic homology groups	Cancer Chromosomes: cytogenetic databases
Journals: detailed information about the journals indexed in PubMed and other Entrez databases	MeSH: detailed information about NLM's controlled vocabulary

Enter terms and click 'GO' to run the search against ALL the databases, OR
Click Database Name or icon to go directly to the Search Page for that database, OR
Click Question Mark for a short explanation of that database.

3. NCBI's Zebrafish Genome Resources

Links to genome-specific resource pages can be found on the Genomic Biology page (<http://www.ncbi.nlm.nih.gov/Genomes/index.html>). A link to the Genomic Biology page is located in the left-hand blue side bar of NCBI's home page.

The Zebrafish Genome Resources page

(<http://www.ncbi.nlm.nih.gov/genome/guide/zebrafish/index.html>)

was created to provide a gateway to web resources and includes links to a number of NCBI sites and to external resources that may be of interest to members of the zebrafish community. The Zebrafish Genome Resources page also includes links to pre-set Entrez queries, including: papers or sequences submitted to NCBI's PubMed or GenBank databases in the last 30 days or the current set of Zebrafish Gene Collection (ZGC) full-length cDNA clones in GenBank.

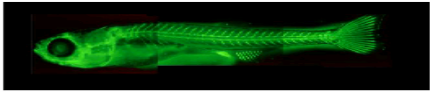
NCBI Home > Genomic Biology > Zebrafish
Genome Resources

Search Entrez for Go

Zebrafish Genome Resources

NCBI Web Resources
The NCBI Handbook.
An online guide to NCBI resources.
BLAST Traces. Search Trace Archive with MegaBLAST.
BLAST Zebrafish. BLAST your sequences against Zebrafish mRNA, EST, and genome sequences.
Clone Registry. Find information about specific clones selected for genome sequencing.
dbSNP. Database of SNPs and other genetic variation.
Entrez Gene. Focal point for genes and associated information.
e-PCR. Review in genomic context your sequence for sequence tagged sites (STSS).
FLcDNA Projects. View Full-Length cDNA Project summaries and links to resources.
GeneRIF. Gene References into Function.
HomoloGene. View calculated and curated homologs for plant and animal genes.
Map Viewer. View and navigate among NCBI's suite of genome-specific Map Viewer and BLAST

Welcome to the **Zebrafish** Genome resources page. This homepage will bring together information on NCBI and community zebrafish-related resources including sequence, mapping, and clone information.



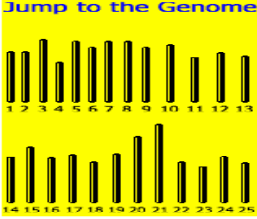
Bone Growth Visualization
Du et al. (2001)
Visualizing Normal and Defective Bone Development in Zebrafish Embryos Using the Fluorescent Chromophore Clacain.
Developmental Biology 238:239-246.
Courtesy of:
The Zebrafish CD Exchange Project's: The Living Laboratory

This month in...
PubMed
GenBank

Upcoming Meeting:
2004 Zebrafish Development and Genetics Meeting
Madison, WI
July 29 - August 2

What's New
April 2004
MGH Primer Selection Tool
Identify simple sequence repeat (SSR)

Jump to the Genome



Community Resources

- ZFIN
- Zebrafish Gene Collection (ZGC)
- Trans-NIH Zebrafish Initiative
- WashU-Zebrafish Genome Resources Project
- Zebrafish International Resource Center
- Tübingen Zebrafish Stockcenter

Maps and Sequencing

- The *Danio rerio* Sequencing Project
- Zebrafish Genome Fingerprinting Project

4. *Danio rerio* Taxonomy Page

Entrez's Taxonomy page for *Danio rerio*

(<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=7955&lvl=3&srchmode=1>) provides another entry point to zebrafish data at NCBI. From the

Taxonomy home page (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=taxonomy>), choose the *Danio rerio* (zebrafish) link to view the *Danio rerio* Taxonomy page.

This page includes a table reporting the current number of zebrafish entries in each Entrez database. The number of entries is linked to the individual databases. Additionally, the *Danio rerio* Taxonomy page includes the list of centers submitting Trace Archive records. (Trace Archive:

<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?>) For each type of **Trace Archive** record

(Clone end, EST, Finishing, Shotgun, WGS, and ALL), view the number of records and follow links, by type, to the Trace Archive. The *Danio rerio* Taxonomy page also includes a list of links to external resources (**LinkOuts**).

NCBI Taxonomy Browser

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search for as complete name ☐ lock

Display levels using filter:

Danio rerio

Taxonomy ID: 7955
 Genbank common name: zebrafish
 Rank: species
 Genetic code: [Translation table 1 \(Standard\)](#)
 Mitochondrial genetic code: [Translation table 2 \(Vertebrate Mitochondrial\)](#)
 Other names:
 synonym: **Brachydanio rerio**
 common name: **zebra fish**
 common name: **zebra danio**

Lineage(full)
 cellular organisms; Eukaryota; Fungi/Metazoa group;
 Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia;
 Chordata; Craniata; Vertebrata; Gnathostomata;
 Teleostomi; Euteleostomi; Actinopterygii; Actinopteri;
 Neopterygii; Teleostei; Elopoccephala; Clupeocephala;
 Otocephala; Ostariophysi; Otophysi; Cypriniphysi;
 Cypriniformes; Cyprinoidea; Cyprinidae; Rasbora; Danio

Entrez records	
Database name	Direct links
Nucleotide	719,379
Protein	14,975
Structure	2
Genome	1
Popset	7
SNP	1,956
3D Domains	2
Domains	1
UniGene	20,978
UniSTS	26,923
PubMed Central	216
Gene	14,482
HomoLoGene	7,198
Taxonomy	1

5. NCBI's FTP resources

In addition to NCBI's web resources, NCBI also provides unrestricted access to NCBI's software and genome data. **NCBI's FTP home page:**

(<http://www.ncbi.nlm.nih.gov/Ftp/index.html>) or directly from NCBI's ftp: <ftp.ncbi.nih.gov>.


6. NCBI's Site Search

A site search (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=ncbisearch>) of NCBI can be conducted either at NCBI's home page by selecting 'NCBI Web Site' from the pull down menu in the top search bar found on most NCBI web pages. Queries submitted from the Entrez home page includes a search of NCBI's web pages.

2. How do I find my gene ?

1. How do I find my gene by searches based on text, sequence or map position ?

Text searches submitted to the **Entrez Gene** database (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>) can include any gene attribute including: symbol, name, alias, identifier (e.g. ZFIN id, GeneID), accession, publication (PubMed ID), GO terms, linkage group or chromosome.



Entrez Gene

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books OMIM

Search for ☒ current records only

Limits Preview/Index History Clipboard Details

Entrez
SITE MAP
Entrez Help

Gene
Search
Gene Help

FAQ

FTP site

Related sites
Entrez Genome
Genomic Biology
HomoloGene
LocusLink
Map Viewer
OMIM
RefSeq
UniGene

Feedback
Help Desk
Corrections
About GeneRIFs

Subscriptions
RefSeq
Gene
Map Viewer

- Enter one or more search terms.
- More information about available fields is available [here](#).
- Consider use of the limits and preview/index functions.
- Remember, boolean operators (AND, OR, NOT) must be in uppercase.

Gene

Background

Gene provides a unified query environment for genes defined by sequence and/or in NCBI's Map Viewer. You can query on names, symbols, accessions, publications, GO terms, chromosome numbers, E.C. numbers, and many other attributes associated with genes and the products they encode.

Because Gene is now an Entrez database, all the familiar and useful functions are now available, including Preview/Index, History, and LinkOut.

Please note: Entrez Gene is under active development. We welcome your [suggestions](#). We have also added a choice in our Feedback/Corrections [form](#) for suggestions; both paths reach the same staff.

Getting started

Sample queries

Look for genes by name part and multiple species [transporter AND \("Drosophila melanogaster"\[orgn\] OR "Mus musculus"\[orgn\]\)](#) [more...](#)


Look for genes by chromosome and symbol [\(11\[chr\] OR 2\[chr\]\) AND adh*\[sym\]](#) [more...](#)

What's new?

June-July, 2004 GeneRIF submissions are supported from Gene. A new [page](#) summarizing the status of the LocusLink->Gene transition is provided.

There are now two types of links between Gene and dbSNP. The one called **SNP** has been provided from the beginning; it results in a listing of all the reference SNPs for a gene. The newer one, **GeneView in dbSNP** connects to the gene-based report of variation associated with a Gene. This report includes a graphical representation of the position of each variation, whether a variant affects a protein sequence, and tables of attributes of all the variants.

Search for a gene of interest by submitting a BLAST search by choosing the type of search from the **BLAST home page** (<http://www.ncbi.nlm.nih.gov/BLAST>).



[BLAST](#)

[PubMed](#)
[Entrez](#)
[BLAST](#)
[OMIM](#)
[Taxonomy](#)
[Structure](#)

NEW 12 May 2004 BLAST 2.2.9 has been released. [Read more...](#)

Info

- FAQs
- News
- References
- NCBI
- Contributors

Education

- Program selection guide
- Tutorial
- URL API guide

Download

- Databases
- Documentation
- Executables
- Source code

Support

- Helpdesk
- Mailing list

Nucleotide

- Discontiguous megablast
- Megablast
- Nucleotide-nucleotide BLAST (blastn)
- Search for short, nearly exact matches
- Search trace archives with megablast or discontiguous megablast

Protein

- Protein-protein BLAST (blastp)
- PHI- and PSI-BLAST
- Search for short, nearly exact matches
- Search the conserved domain database (rpsblast)
- Search by domain architecture (cdart)

Translated

- Translated query vs. protein database (blastx)
- Protein query vs. translated database (tblastn)
- Translated query vs. translated database (tblastx)

Genomes

- Chicken, cow, pig, dog, sheep, cat **NEW**
- Environmental samples
- Human, mouse, rat
- Fugu rubripes, zebrafish
- Insects, nematodes, plants, fungi, malaria
- Microbial genomes, other eukaryotic genomes

Special

- Search for gene expression data (GEO BLAST)
- Align two sequences (bl2seq)
- Screen for vector contamination (VecScreen)
- Immunoglobulin BLAST (IgBlast)

Meta

- Retrieve results by RID
- Get this page with javascript-free links

BLAST tools enable searches against nucleotide or protein databases, translational searches (protein vs. nucleotide), **MegaBLAST** comparisons, genome specific BLAST pages (including a **zebrafish BLAST page**: <http://www.ncbi.nlm.nih.gov/genome/seq/DrBlast.html>) and BLASTing of two nucleotide or protein sequences via bl2seq.


NCBI Genomic Biology Zebrafish Genome Resources

Search for

BLAST
[overview](#)
[FAQs](#)
[news](#)
[manual](#)
[references](#)

MegaBLAST
[Danio Traces](#)
 enter a nucleotide query sequence and get a match, see a tblastx alignment of your two sequences

Blast The Zebrafish Genome


Danio rerio
 Blast your sequence against zebrafish specific sequences.

Database: Program:
☒ use

 EST
 HTGS
 WGS Traces
 EST Traces
 Reference mRNAs
 Reference Proteins

Enter an accession number or sequence in FASTA format:

Optional parameters
[Expect](#) [Filter](#) [Descriptions](#) [Alignments](#)

[Advanced options:](#)

The “Options for advanced blasting” section on the BLAST query page allows searches for hits against two organisms with the “AND” option or to search for hits in either organism with the “OR” option such as: *Xenopus tropicalis* OR *Takifugu rubripes*.

Alternatively, view the pre-computed nucleotide and protein sequence comparisons available in NCBI's **HomoloGene**

(<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene>) and BLink

(e.g. <http://www.ncbi.nlm.nih.gov/sutils/blink.cgi?pid=42542497>) resources.

Identify a gene of interest by its map position. Search by linkage group or chromosome in **Entrez Gene** or query map data by starting at NCBI's **Map Viewer** home page (<http://www.ncbi.nlm.nih.gov/mapview/>) or the zebrafish genome view page (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=7955).

NCBI **NCBI Map Viewer**

Genome Taxonomy Entrez BLAST Help

Search for

New! - searching for map objects in any user-determined subset of all plant genomes presented by NCBI. A map object includes, but is not limited to, a locus, probe name, GenBank accession, gene or name of BAC clone. Select "All plants" from the search menu at the top of the page or click (S) to the right of the Plants node.

Click the **B** to BLAST
Click the **S** to search the group
Click on the binomial to view the genome overview

Other Vertebrates
B *Danio rerio* (zebrafish)

Mammals
B *Homo sapiens* (human)
B *Mus musculus* (mouse)
B *Rattus norvegicus* (rat)
B *Bos taurus* (cow)
B *Sus scrofa* (pig)
B *Canis familiaris* (dog)
B *Felis catus* (cat)
B *Ovis aries* (sheep)

Invertebrates
B *Anopheles gambiae* (mosquito)
B *Caenorhabditis elegans* (nematode)
B *Drosophila melanogaster* (fruit fly)
B *Apis mellifera* (honey bee)

Fungi
B *Saccharomyces cerevisiae* (baker's yeast)
B *Schizosaccharomyces pombe* (fission yeast)
B *Neurospora crassa*
B *Magnaporthe grisea*
B *Eremothecium gossypii*
B *Encephalitozoon cuniculi*

Protozoa
B *Plasmodium falciparum*

Plants **B** **S**
B *Arabidopsis thaliana* (thale cress)
B *Avena sativa* (oat)
B *Hordeum vulgare* (barley)
B *Oryza sativa* (rice)
B *Triticum aestivum* (wheat)
B *Zea mays* (corn)
B *Lycopersicon esculentum* (tomato)
B *Glycine max* (soybean)

See more about **B** *Bacteria*, **B** *Organelles*, **B** *Viruses*

The Map Viewer supports search and display of genomic information by chromosomal position. Regions of interest can be retrieved by text queries (e.g. gene or marker name) or by sequence alignment (BLAST). View results at the whole genome level, and select what to display in more detail. Multiple options exist to configure your display, download data, navigate to related data, and analyze supporting information using the tools provided. [More...](#)

Search zebrafish map data at the Map Viewer home page by selecting an organism from the Search drop down menu and entering a gene symbol, accession or alias to the query bar at the top of the page. Other terms may be used for searching, depending on the data associated with each organism.

Following the link to any of the organism-specific genome view pages (from the Map Viewer home page), submit a search in the top menu bar or click on the “Advanced Search” button to further specify their query. For zebrafish, the search can be refined by linkage group, marker type (SNP or non-SNP) or the map name.



Map data can also viewed by choosing any of the LG links in the “Jump to the Genome” figure on the **Zebrafish Genome Resources** page (<http://www.ncbi.nlm.nih.gov/genome/guide/zebrafish>). These links go directly to each linkage group in the zebrafish Map Viewer.

2. How do I find my gene when a text search does not work ?

One approach is to take advantage of the higher level of annotation available for other organisms to find a gene of interest. Although a gene may not yet be identified in one organism, it may be in another. Even though a gene of interest may not yet be identified, in zebrafish for example, mRNA or protein sequences representing that gene may be available either in the sequence databases (**Entrez Nucleotide** or **Entrez Protein**) or in an **Entrez Gene** record for an uncharacterized gene.

To begin, search Entrez Gene by the gene name or gene symbol of interest. If there is a matching Entrez Gene record, follow the links to related records provided in the Links pull-down menu to look for related mRNAs or proteins representing the zebrafish gene.

Following links to **Homologene** to view other highly related mRNAs and proteins.



Homologene
Discover Homologs
[Help](#)

[Entrez](#)
[PubMed](#)
[Nucleotide](#)
[Protein](#)
[Genome](#)
[Structure](#)
[Map V](#)

Search for

[Limits](#)
[Preview/Index](#)
[History](#)
[Clipboard](#)
[Details](#)

About Entrez

HomoloGene

Home

Query Tips

Build Procedure

FTP Site

Genome Resources

Homo sapiens

Mus musculus

Rattus norvegicus

Danio rerio

HomoloGene is a system for automated detection of homologs among the annotated genes of several completely sequenced eukaryotic genomes.

HomoloGene Release Statistics

Initial numbers of genes from complete genomes, numbers of genes placed in a homology group, and the numbers of groups for each species

HomoloGene Build

36

Species	Number of genes		HomoloGene groups
	Input	Grouped	
H.sapiens	22,827	18,055	16,782
M.musculus	24,019	19,996	18,036
R.norvegicus	20,913	17,429	16,042
D.melanogaster	12,918	8,717	7,683
A.gambiae	12,012	8,543	7,577
C.elegans	19,109	6,502	5,260
S.pombe	4,947	3,625	3,359
S.cerevisiae	5,863	3,612	3,146
N.crassa	10,079	6,156	6,049
M.grisea	11,109	6,307	6,028
A.thaliana	26,281	8,022	4,791
P.falciparum	5,222	1,770	1,589

Last updated on: 05/25/2004

Related Resources

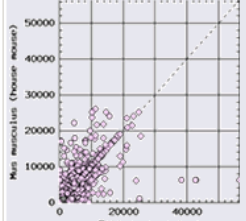
Entrez Genome

A collection of complete genome sequences that includes more than 1000 viruses and over 100 microbes

- Archaea
- Bacteria
- Eukaryota
- Viruses

Tax Plot

Three-way view of genome similarities



Each Homologene record also includes related **UniGene** clusters. Examine the sequences included in the UniGene cluster by following the link to the UniGene database. If an Entrez Gene record is associated with the UniGene cluster, a link to Gene will be provided in the UniGene Links menu.

NCBI **UniGene**

Entrez PubMed Nucleotide Protein Genome

Search UniGene for Go Clear

Limits Preview/Index History Clipboard Details

NCBI

UniGene
 Homepage
 Query Tips
 FAQ
 DDD
 Download
 UniGene

Related Resources
 LocusLink
 HomoloGene
 dbEST
 Trace Archive
 BLAST
 CGAP

UniGene is an experimental system for automatically partitioning GenBank sequences into a non-redundant set of gene-oriented clusters. Each UniGene cluster contains sequences that represent a unique gene, as well as related information such as the tissue types in which the gene has been expressed and map location.

Species Entries	UniGene
Chordata	
Mammalia	
Bos taurus	24,195
Canis familiaris	15,665
Homo sapiens	106,934
Mus musculus	76,298
Ovis aries	3,160
Rattus norvegicus	39,050
Sus scrofa	21,947
Aves	
Gallus gallus	20,155
Amphibia	
Xenopus laevis	23,754
Xenopus tropicalis	14,632
Actinopterygii	
Danio rerio	20,978
Oncorhynchus mykiss	14,178
Oryzias latipes	8,133
Salmo salar	1,076
Ascidiacea	
Ciona intestinalis	14,098
Echinodermata	
Echinoidea	

Alternatively, follow the Protein link to **Entrez Protein** to view related proteins by following the **BLink (BLAST Link)** associated with the Protein record. BLink provides a graphical display to view the pre-computed protein comparisons between the query protein and all other proteins in Entrez Protein. These results may identify a zebrafish protein that is highly similar to the query protein. Check the Links associated with the Entrez Protein record to determine if that protein is associated with an Entrez Gene record.

Address <http://www.ncbi.nlm.nih.gov/subs/blink.cgi?pid=5174681&all=1>

NCBI

BLAST Protein Structure PubMed Taxonomy
Genome Nucleotide 3D-Domains Books Help

Query: gi|5174681 sine oculis homeobox homolog 1 [Homo sapiens]
Matching gi: 1246761, 2495290, 12744793

Best hits Common Tree Taxonomy Report 3D structures CDD-Search GI list

200 BLAST hits to 22 unique species [Sort by taxonomy proximity](#)

0 Archaea 0 Bacteria 197 Metazoa 2 Fungi 0 Plants 0 Viruses 0 Other Eukaryotae

Keep only Cut-Off 100 Select Reset

284 aa

	SCORE	P	ACCESSION	GI	PROTEIN DESCRIPTION
	1509	27	CAA62974	1246761	six1 [Homo sapiens]
	1509	27	Q15475	2495290	Homeobox protein SIX1 (sine oculis homeobox homolog 1)
	1509	27	AAK06772	12744793	SIX1 [Homo sapiens]
	1501	27	AAH08874	14250808	Sine oculis homeobox homolog 1 [Homo sapiens]
	1501	27	AAP35746	30583005	sine oculis homeobox homolog 1 (Drosophila) [Homo sapiens]
	1501	1	AAP36799	30585053	Homo sapiens sine oculis homeobox homolog 1 (Drosophila) [Homo sapiens]
	1498	21	AAH23304	23273881	Sine oculis homeobox homolog 1 [Mus musculus]
	1498	21	XP_343080	34865437	sine oculis homeobox homolog 1 [Rattus norvegicus]
	1492	21	NP_033215	33859494	sine oculis homeobox homolog 1 [Mus musculus]
	1441	21	AAK11607	13022007	homeobox protein SIX1 [Rattus norvegicus]
	1429	21	CAA56585	791084	six1 [Mus musculus]
	1429	21	S60751	2137402	homeotic protein six1 - mouse (fragment)
	1419	17	AAF91422	9652166	homeobox protein SIX1 [Xenopus laevis]
	1398	15	AAH66396	42542497	Hypothetical protein zgc:77345 [Danio rerio]
	1390	15	CAG10546	47230132	unnamed protein product [Tetraodon nigroviridis]
	1124	15	NP_571858	18859361	sine oculis homeobox homolog 2.1 [Danio rerio]
	1123	21	Q62232	2851595	Homeobox protein SIX2 (sine oculis homeobox homolog 2)

3. How can I find the zebrafish homolog of a recently cloned mouse gene?

If searches of the pre-computed comparisons in HomoloGene or BLink do not yield any likely candidates, another approach would be to submit a search of the mouse gene symbol to Entrez's home page to determine if a similarly named zebrafish gene exists.

Alternatively, Homologene comparisons may have already identified a UniGene cluster containing ESTs or uncharacterized cDNAs that represent the zebrafish

homolog. If a ZFIN marker or Entrez Gene record contains these uncharacterized sequences, a link to ZFIN or Entrez Gene will be at the top of the UniGene page.

3. How do I find gene-related information ?

1. What gene-related information is available in Entrez Gene ?

The screenshot displays the NCBI Entrez Gene web interface. The top navigation bar includes links for Entrez, PubMed, Nucleotide, Protein, Genome, Structure, PMC, Taxonomy, Books, and OMIM. The main search area shows a search for 'Gene' with options for Limits, Preview/Index, History, Clipboard, and Details. The search results for 'zgc:77345' in 'Danio rerio' are displayed, including the GeneID (404627), Locus tag (ZDB-GENE-040426-2308), Gene type (protein coding), Gene name (zgc:77345), Gene description (zgc:77345), RefSeq status (Predicted), and Organism (Danio rerio). The Lineage is listed as Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio. The Gene aliases are MGC77345. The Bibliography section shows Gene References into Function (GeneRIF) with links to Submit and help. The General gene information section includes Markers (Sequence Tagged Sites/STS) with a link to fc88a09.x1 (e-PCR). The General protein information section shows the Name as hypothetical protein zgc:77345. The NCBI Reference Sequences (RefSeq) section lists the mRNA Sequence (NM_207095), Source Sequence (BC066396), and Product (NP_996978 hypothetical protein zgc:77345).

Entrez Gene records include gene symbols, gene names, links to external data providers (e.g. ZFIN, MGI, HGNC), aliases, related sequences (including RefSeqs (<http://www.ncbi.nlm.nih.gov/RefSeq/>), map location, taxonomy information and links to publications (PubMed), marker information linked to **UniSTS** (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists>). Links to other related

resources are included at the bottom of the Entrez Gene page and in the Links pull-down menu.

NCBI UniSTS Integrating Markers and Maps

Search UniSTS for [] Go Clear

Limits Preview/Index History Clipboard Details

UniSTS is a comprehensive database of sequence tagged sites (STSs) derived from STS-based maps and other experiments. STSs are defined by PCR primer pairs and are associated with additional information such as genomic position, genes, and sequences.

Summary of Maps in UniSTS

Organism	Map name	Map title	Total markers	Total UniSTSs ¹	Links ²
Apis mellifera	Hunt	Hunt Honeybee Linkage Map 3_10_04	1,055	38	
	Solignac	Solignac Genetic Map	1,096	688	
Bos taurus	ILTX-2004	5000 Rad Cattle RH Map - 2004	1,911	1,873	
	MARC	MARC Bovine Linkage Map	1,419	1,386	
Canis familiaris	RHDF5000	Canine 1Mb RH Map	3,092	3,037	
Danio rerio	GAT	Gates Haploid Panel	422	246	
	HS	Heat Shock Diploid Cross	6,017	722	
	LN54	Loeb/NIH/5000/4000	4,595	817	
	MGH	Boston MGH Cross	3,855	3,425	
	MOP	Mother of Pearl	713	347	
	T51	Goodfellow T51	15,306	2,019	
	ZMAP	Zebrafish Genome Integrated Map	25,920	5,081	
Felis catus	Genetic	Genetic linkage map of the domestic cat	328	320	
	RH	Domestic cat radiation hybrid map	1,126	1,118	
Gallus gallus	WURC	Wageningen University Sheep linkage map	2,110	973	
Homo sapiens	GM99-G3	Human Transcript Map 99	7,056	7,056	

View suggested query options by clicking on the Limits tab at the top of the Entrez Gene page (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=gene>).

Subscriptions to announcement lists for changes to the RefSeq, Entrez Gene and Map Viewer projects are available. Links to the Subscription information pages are found at the bottom of the blue side bar in Entrez Gene.

RefSeq announce: <http://www.ncbi.nlm.nih.gov/mailman/listinfo/refseq-announce>

Gene announce: <http://www.ncbi.nlm.nih.gov/mailman/listinfo/gene-announce>

Map Viewer announce: <http://www.ncbi.nlm.nih.gov/mailman/listinfo/mapview-announce>

2. How do I identify a likely homolog for my gene ?

Highly related nucleotide and protein sequences representing putative homologs between multiple genomes are identified through computational analysis and presented in the Homologene database.

Submit a text based query to HomoloGene from the Homologene home page, (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene>), the Entrez home page or by choosing Homologene from any of the Search bars found at the top of most NCBI pages.

Links to Homologene are also provided in the drop down Links menu of the other Entrez databases including Entrez Gene, UniGene, Entrez Nucleotide, Entrez Protein, Taxonomy, GEO and PubMed.

3. Is my gene annotated on the zebrafish whole genome assembly ?

As NCBI has done for projects, the zebrafish genome sequence will become available in NCBI's Map Viewer when the zebrafish assembly sequences are accessioned in GenBank/DDBJ/EMBL. At that time, the zebrafish genome sequence will be added to the set of organisms included in NCBI's genome pipeline.

4. What map data is available for zebrafish at NCBI ?

The Zebrafish Map Viewer (<http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?taxid=7955>) at NCBI includes the LN54 and T51 radiation hybrid maps, the genetic maps (MGH, HS, MOP, GAT), the Zebrafish Genome Integrated Map (ZMAP), and the SNP (Single Nucleotide Polymorphism) variation map.

5. What other genes, ESTs, STS markers or BACs are annotated near my gene ?

Search NCBI's **Map Viewer** to identify candidate genes or markers in a region of interest. View marker placement on the genetic and radiation hybrid maps by submitting a text query.

Identify a BAC clone that contains a marker or gene of interest or that is located in a specific genomic region by submitting a search to NCBI's Clone Finder (<http://www.ncbi.nlm.nih.gov/genome/clone/clonefinder/CloneFinder.html>). At this time, Clone Finder searches are limited to mouse.

6. How can I identify BAC and PAC clones that would aid me in mapping genomic clones to chromosomes?

Find all mapped PACs or BAC_ENDs in NCBI's Map Viewer by submitting a search based on the naming schema for zebrafish PACs and BACs. Search PACs with busm1* and BAC_ENDs with bz*.

NCBI's clone registry of genomic clones (<http://www.ncbi.nlm.nih.gov/genome//clone/>) can be queried to view sequence, library, map, and distributor information. View

library information for each organism in the Library Browser

(<http://www.ncbi.nlm.nih.gov/genome//clone/clbrowse.cgi>) and links to clone distributors on the Distributors page (<http://www.ncbi.nlm.nih.gov/genome//clone/distributors.html>). Follow the pre-computed links associated with the nucleotide accession found on the Clone page to mine the data stored in other NCBI databases including UniSTS, UniGene, Map Viewer (for example:

<http://www.ncbi.nlm.nih.gov/genome/clone/clname.cgi?stype=ld&list=1145501&TransHist=0>).

7. How do I find all ZGC clones ?

Download the list of full-length clones (<http://zgc.nci.nih.gov>) by clicking on the Full Text button to view the Full length clone list or FASTA sequences

(<ftp://ftp1.nci.nih.gov/pub/MGC/fasta/>) from the ZGC web site (<http://zgc.nci.nih.gov/>).


Another option is to submit a search against the Entrez Nucleotide database:

“Danio rerio [organism] AND biomol_mrna [properties] AND MGC [keyword]” ([http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&db=Nucleotide&orig_db=Nucleotide&Nucleotidefilters=truen&term=Danio+AND+rerio+AND+MGC\[KYWD\]](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&db=Nucleotide&orig_db=Nucleotide&Nucleotidefilters=truen&term=Danio+AND+rerio+AND+MGC[KYWD])).


A short-cut has also been provided for this query on the Zebrafish Genome Resources page (<http://www.ncbi.nlm.nih.gov/genome//guide/zebrafish/>).

8. How do I find library information for a ZGC clone ?

From the ZGC home page (<http://zgc.nci.nih.gov/>), choose the Table or Full Text option in the ZGC Full-length Clone Information table to view the zebrafish library list choose either the Table or Full text button for the ZGC library list on the ZGC home page (<http://zgc.nci.nih.gov/>) .



Zebrafish Gene Collection



ZGC

- [Home](#)
- [Project Summary](#)
- [Collaborators](#)
- [Purchase Clones](#)
- [Vectors and Method Overviews](#)

Sequencing Info

- [Candidate Clones for Genes](#)
- [ZGC ESTs](#)

Other Species Collections

- [MGC](#)
- [XGC](#)

Related Links

- [Trans-NIH Zebrafish Initiative](#)
- [Zebrafish Genome Resources](#)
- [ZFIN](#)

ZEBRAFISH GENE COLLECTION

As of 19-Jul-04	#
Total ZGC full-length ORF clones	4,822
Non-redundant genes	4,215

About the ZGC

The Zebrafish Gene Collection (ZGC) is an NIH initiative that supports the production of cDNA libraries, clones and sequences to provide a complete set of full-length (open reading frame) sequences and cDNA clones of expressed genes for zebrafish. All resources generated by the ZGC are publicly accessible to the biomedical research community. For an overview, see the [ZGC Project Summary](#).

Search for Full-length ZGC Clones by Gene Symbol or Keyword

Enter Gene Symbol

Search

[Help](#)

Enter Gene Keyword

Search

[Help](#)

Nucleotide BLAST against ZGC Clones

[Nucleotide BLAST](#) a sequence against ZGC **full-length** sequences.

ZGC Full-length Clone Information

Full length clone list	Full text	Help
Full length clone sequences	Fasta Format	Help
ZGC library list	Table or Full text	Help

If you have any questions, comments, or need information about ZGC, please contact [Dr. Rebekah Rasooly](#).

9. Are there any SNPs associated with my gene?

Query NCBI's Single Nucleotide Polymorphism (SNP) database

(<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=snp>) (e.g. SNP identifier, rs372494:

http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=3729494).

Identify all zebrafish SNPs known to be associated with an Entrez Gene record, by the submitting the query: snp_gene[filter] AND "Danio rerio"[organism] to dbSNP.

Identify all mapped SNPs in NCBI's by submitting the query: `zsnp*` to the zebrafish genome view page.

The screenshot shows the NCBI Zebrafish genome view interface. At the top, there are tabs for Nucleotide, Protein, Genome, Structure, Popset, Taxonomy, and SNP. The 'SNP' tab is selected. Below the tabs, there is a search bar with the query 'snp_gene[filter] AND "Danio rerio"[organism]' and buttons for 'Go' and 'Clear'. Below the search bar, there are tabs for Limits, Preview/Index, History, Clipboard, and Details. The 'Preview/Index' tab is selected. Below the tabs, there is a 'Display' dropdown set to 'Graphic Summary', a 'Show' dropdown set to '20', a 'Sort' dropdown, and a 'Send to' dropdown set to 'Text'. Below this, it says 'Items 1-20 of 606' and 'Page 1 of 31 Next'. The main content area shows four search results, each with a checkbox, a link to the SNP (rs3729398, rs3729394, rs3729389, rs3729379), the organism name 'Danio rerio', and a 'Links' button. Each result also has a small graphic showing the SNP location on a chromosome.

10. What Entrez Gene records have links to the ZFIN's expression data ?

Entrez Gene records contain links to the Gene Expression Database (GXD) at ZFIN in the drop down 'Links' menu in the upper right corner of the Entrez Gene page, for example: <http://zfin.org/cgi-bin/webdriver?MIval=aa-xpatview.apg&OID=ZDB-XPAT-020319-2>.

Retrieve the list of all Entrez Generecords containing a GXD link with the query: "Danio rerio"[organism] AND ZFIN[Text Word] AND Expression[Text Word].

11. What resources are available to find expression data ?

Expression-related records in any of the Entrez databases can be found by submitting a query from the Entrez home page (<http://www.ncbi.nlm.nih.gov/Entrez/index.html>) (e.g. notochord AND development AND "Danio rerio"[organism]).

Additional expression-related resources include Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) to view GEO datasets and profiles. The GEO database can be queried by gene name, organism or tissue, and provide an additional source of functional information for multiple organisms. UniGene's DDD (Digital Differential Display, (http://www.ncbi.nlm.nih.gov/UniGene/info_ddd.html) provides gene expression fingerprints for cancer cells and their normal counterparts and UniGene's Library Browser (<http://www.ncbi.nlm.nih.gov/UniGene/lbrowse2.cgi?TAXID=7955>) includes the number of sequences and clusters associated with each library and detailed library information.

12. What are other possible ways to infer function for my gene?

Entrez Gene records may include submissions to the Gene References into Function (GeneRIF) database. GeneRIF submissions include short descriptions of a gene's function and the associated publication. Information on the GeneRIF project can be found at: <http://www.ncbi.nlm.nih.gov/projects/GeneRIF/GeneRIFhelp.html>.

GeneRIFs can be submitted by the research community from any Entrez Gene page by clicking on the Submit link found in the Bibliography gray bar on the Entrez Gene details page (e.g.

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&dopt=Graphics&list_uids=30483).

Conserved domains may also provide functional information. The Conserved Domain database (CDD) can be queried by a text search (e.g. pham00757, smart00261, cd00064 or Furin-like) from the Entrez home page

(<http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>) or directly from the CDD home page (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=cdd>) to domain details and an alignment of proteins containing that domain.

Conserved domains can be identified by submitting a protein sequence (accession, Gi or Sequence in FASTA format) to CDD's CD-Search form (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

Functional information can also be inferred from homologous sequences identified in NCBI's Homologene database (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Submitting the query "Danio rerio"[organism] against the Homologene database identifies all Homologene sets that contain a related zebrafish UniGene cluster.

13. How do I find cDNAs that have been derived from a specific library?

One approach would be to go to the Zebrafish Gene Collection home page (<http://zgc.nci.nih.gov/>), under the heading "ZGC Full-length Clone Information"

select the “Table” or “Full Text” buttons to view the ZGC library list details (Library, Tissue, Vector or number of Clones). From the ZGC home page (<http://zgc.nci.nih.gov>), navigate to the ZGC library list page, choose the library name to go to the Library Info Page or the Clones link to view the list of IMAGE ids, GenBank accessions, symbols and the GenBank definition that correspond to each sequenced clone.

Another approach would be to begin in UniGene, navigate to the details for each EST by clicking on the EST accession number to view the Sequence Information page. View the Library Description, the Sequence Submitters, the number of UniGene clusters containing members of the library or the number of sequences in each cluster by clicking on the linked Library ID. The list of cDNAs for each cluster can be seen by then clicking on the UniGene cluster number.

Example: zebrafish shield stage whole embryo (dbEST Library ID.9774)

<http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?ORG=Dr&LID=9774> .

14. Where can I find a list of all Entrez records for *Danio rerio* ?

NCBI’s Taxonomy database includes an index of the records included in all Entrez databases for each organism represented by at least one sequence in either the Entrez Nucleotide or the Entrez Protein database.

From the Taxonomy home page

(<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Taxonomy>)

navigate to the zebrafish-specific page

(<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=7955&lvl=3&srchmode=1>) by clicking on the Danio rerio (zebrafish) link.

The **Taxonomy database** page also includes links to each linkage group in Map Viewer and a table of Trace (raw DNA single-pass reads) record counts from each sequencing center, divided by the type of Trace (for example: CLONEEND, EST, FINISHING, SHOTGUN, WGS or ALL).

Further down the Taxonomy database page links are provided to external resources pertinent to that organism. These links are called LinkOuts. Detailed information on creating LinkOuts at NCBI can be found on the **LinkOut** home page (<http://www.ncbi.nlm.nih.gov/entrez/linkout/>).

Near the bottom of the Taxonomy database page, the “Show organism modifiers” link provide additional links to subsets of nucleotide (blue number) and protein (red number) sequences based on their strain names, isolate name, or specimen-voucher name. For example, clicking on the blue SJD link returns the set of nucleotide records sequenced from the SJD strain

([http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=nucleotide&cmd=Search&dopt=DocSum&term=txid7955\[orgn\]+AND+strain+SJD\[tw\]](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=nucleotide&cmd=Search&dopt=DocSum&term=txid7955[orgn]+AND+strain+SJD[tw])).

4. Where can I ask questions or submit suggestions ?

NCBI welcomes questions and suggestions.

A link to contact information is provided at the bottom of the blue side bar on the NCBI home page (http://www.ncbi.nlm.nih.gov/About/glance/contact_info.html).

Submit questions directly to NCBI's Help Desk at info@ncbi.nlm.nih.gov.

Questions or suggestions for NCBI's RefSeq, Entrez Gene or OMIM databases can also be submitted via the Corrections form

(<http://www.ncbi.nlm.nih.gov/RefSeq/update.cgi>). A link to this form is provided in the blue side bar of the Entrez Gene pages.